View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Elsevier - Publisher Connector

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Theoretical Computer Science 355 (2006) 243–260

Theoretical Computer Science

www.elsevier.com/locate/tcs

Probability distribution for simple tautologies

Marek Zaionc*, [1]

Computer Science Department, Jagiellonian University, Nawojki 11, 30-072 Kraków, Poland

Abstract

In this paper we investigate the size of the fraction of tautologies of the given length n against the number of all formulas of length n for implicational logic. We are specially interested in asymptotic behavior of this fraction. We demonstrate the relation between a number of premises of implicational formula and asymptotic probability of finding formula with this number of premises. Furthermore, we investigate the distribution of this asymptotic probabilities. Distribution for all formulas is contrasted with the same distribution for tautologies only. We prove those distributions to be so different that enable us to estimate likelihood of truth for a given long formula. Despite the fact that all discussed problems and methods in this paper are solved by mathematical means, the paper may have some philosophical impact on the understanding how much the phenomenon of truth is sporadic or frequent in random logical sentences.

© 2006 Elsevier B.V. All rights reserved.

Keywords: Asymptotic probability in logic

1. Introduction

Probabilistic methods appear to be very powerful in combinatorics and computer science. A point of view of those methods is that we investigate the typical object chosen from the set. In this paper we investigate the lower bound of the proportion between the number of formulas of the size n that are tautologies against the number of all formulas of size n for propositional formulas. Our interest lays in finding limit of that fraction when $n \to \infty$. If the limit exists it represents the real number between 0 and 1 which we may call *the density of truth* for the logic investigated. After isolating the special class of formulas called simple tautologies we prove that their fractions among all formulas converges. We conjecture that indeed the fraction of tautologies, for large k, is very close to the lower bound determined by simple tautologies. In general we are interested in finding the *density* of some special subclasses of formulas. This paper is a part of the research in which the likelihood of truth for the given propositional logic with a restricted number of variables is estimated. Consult for example paper of Moczurad et al. [4] for purely implicational logic of one variable (and at the same time a type system) and Zaionc [8] for the classical logic of implication and negation. In the paper of Kostrzycka and Zaionc [3] the exact proportion between intuitionistic and classical logics of the same language have been found. Compare also two papers of Dershowitz and Harris [2] and Harris [2] where asymptotic probability of

* Tel.: +48 12 664 6783; fax: +48 12 634 1865.
E-mail address: zaionc@ii.uj.edu.pl.

[1] Supported by the State Committee for Scientific Research in Poland (KBN), research Grant 7T11C 022 21.

[2] N. Dershowitz, M. Harris, Enumerating the propositional formulas equivalent to a Boolean function, private communication, see web page http://www.math.tau.ac.il/ nachumd/

0304-3975/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.tcs.2006.01.002

satisfiability of propositional formulas is considered. All papers cited above describes asymptotic results in the logical systems with a restricted number of variables.

In this paper we investigate the language $\mathcal{F}_k$ consisting of implicational formulas over k propositional variables.

Definition 1. The language $\mathcal{F}_k$ over k propositional variables $\{a_1, \ldots, a_k\}$ is defined inductively as

$a_i \in \mathcal{F}_k \quad \forall i \leqslant k$
$\phi \to \psi \in \mathcal{F}_k$ if $\phi \in \mathcal{F}_k$ and $\psi \in \mathcal{F}_k$

First we have to establish the way the size of formulas are measured.

Definition 2. By $\|\phi\|$ we mean the size of formula ϕ which we define as the total number of occurrences of propositional variables in the formula. Parentheses which are sometimes necessary and the implication sign itself are not included in the size of formula. Formally,

$$\|a_i\| = 1 \text{ and } \|\phi \to \psi\| = \|\phi\| + \|\psi\|.$$

Definition 3. We associate the density $\mu(\mathcal{A})$ with a subset $\mathcal{A} \subseteq \mathcal{F}_k$ of formulas as

$$\mu(\mathcal{A}) = \lim_{n\to\infty} \frac{\#\{t \in \mathcal{A} : \|t\| = n\}}{\#\{t \in \mathcal{F}_k : \|t\| = n\}} \tag{1}$$

if the limit exists.

The number $\mu(\mathcal{A})$ if it exists is an asymptotic probability of finding a formula from the class $\mathcal{A}$ among all formulas from $\mathcal{F}_k$ or it can be interpreted as the asymptotic density of the set $\mathcal{A}$ in the set $\mathcal{F}_k$. It can be seen immediately that the density μ is finitely additive so if $\mathcal{A}$ and $\mathcal{B}$ are disjoint classes of formulas such that $\mu(\mathcal{A})$ and $\mu(\mathcal{B})$ exist then $\mu(\mathcal{A} \cup \mathcal{B})$ also exists and $\mu(\mathcal{A} \cup \mathcal{B}) = \mu(\mathcal{A}) + \mu(\mathcal{B})$. It is straightforward to observe that for any finite set $\mathcal{A}$ the density $\mu(\mathcal{A})$ exists and is 0. Dually for co-finite sets $\mathcal{A}$ the density $\mu(\mathcal{A}) = 1$. The density μ is not countably additive so in general the formula

$$\mu\left(\bigcup_{i=0}^{\infty} \mathcal{A}_i\right) = \sum_{i=0}^{\infty} \mu(\mathcal{A}_i) \tag{2}$$

is not true for all pairwise disjoint classes of sets $\{\mathcal{A}_i\}_{i\in\mathbb{N}}$. A good counterexample for Eq. (2) is to take as $\mathcal{A}_i$ the ith formula from our language under any natural order of formulas. On the left hand side of Eq. (2) we get $\mu(\mathcal{F}_k)$ which is 1 but on right hand side $\mu(\mathcal{A}_i) = 0$ for all $i \in \mathbb{N}$ and so the sum is 0.

In this paper we are specially interested in the distribution of densities with respect to some numerical syntactic property of formulas.

Definition 4. By a random variable X we understand the function $X : \mathcal{F}_k \mapsto \mathbb{N}$ which assigns a number $n \in \mathbb{N}$ to the implicational formula in such a way that for any n the density $\mu(\{\phi \in \mathcal{F}_k : X(\phi) = n\})$ exists and moreover

$$\sum_{n=0}^{\infty} \mu(\{\phi \in \mathcal{F}_k : X(\phi) = n\}) = 1.$$

Definition 5. By the distribution of a random variable X we mean the function $\overline{X} : \mathbb{N} \mapsto \mathbb{R}$ defined by

$$\overline{X}(n) = \mu(\{\phi \in \mathcal{F}_k : X(\phi) = n\}).$$

Definition 6. The expected value $E(\overline{X}) = \sum_{p=0}^{\infty} p \cdot \overline{X}(p)$ of distribution X, variance $Var(\overline{X}) = E(\overline{X}^2) - (E(\overline{X}))^2 = \sum_{p=0}^{\infty} p^2 \overline{X}(p) - (E(\overline{X}))^2$ and standard deviation $\sigma(\overline{X}) = \sqrt{Var(\overline{X})}$ are defined in conventional way.

In the paper of Moczurad et al. [4] we showed what is the relation between the number of premises of implicational formula and asymptotic probability of finding a formula with this number of premises. In this paper we are going to

investigate the distribution of densities with respect to the number of premises but only for simple tautologies, which form a large subset of all tautologies. We prove that this distribution is so different from the previous one that it can be used to distinguish a tautology only by counting the number of its premises.

2. Elementary counting of implicational formulas

In this section we present some properties of numbers characterizing the amount of formulas in different classes defined in our language. We may observe that many results and methods could be rephrased purely in terms of binary trees with given properties. Obviously an implicational formula from $\mathcal{F}_k$ of size n can be seen as a binary tree with n leaves and k labels per leave.

Definition 7. By F_n^k we mean the total number of formulas from $\mathcal{F}_k$ of size n so

$$F_n^k = \#\{\phi \in \mathcal{F}_k : \|\phi\| = n\}. \tag{3}$$

Lemma 8. *F_n^k is given by the following recursion*:

$$F_0^k = 0, \quad F_1^k = k, \tag{4}$$

$$F_n^k = \sum_{i=1}^{n-1} F_i^k F_{n-i}^k. \tag{5}$$

Proof. We may use combinatorial observation. Formula from $\mathcal{F}_k$ of size n can be interpreted as full binary tree of n leaves with k label per leaf. Therefore for $n = 0$ and $n = 1$ it is obvious. Any formula of size $n > 1$ is the implication (tree) between some pair of formulas (trees) of sizes i and $n - i$, respectively. Therefore the total number of such pairs is $\sum_{i=1}^{n-1} F_i^k F_{n-i}^k$. □

Lemma 9. *The number $F_n^k = k^n C_n$ where C_n is $(n-1)$th Catalan number.*

Proof. Indeed, numbers C_n are given by similar recursion schema:

$$C_0 = 0 \quad C_1 = 1, \tag{6}$$

$$C_n = \sum_{i=1}^{n-1} C_i C_{n-i}. \tag{7}$$

Therefore by simple induction we can immediately see that $F_n^k = k^n C_n$. Obviously for formulas build with just one propositional variable we have $F_n^1 = C_n$. □

For more elaborate treatment of Catalan numbers see Wilf [7, pp. 43–44]. We mention only the following well-known nonrecursive formula for C_n.

$$C_n = \frac{1}{n}\binom{2n-2}{n-1}, \tag{8}$$

and repeat some simple properties which are its consequences. For every $n \geqslant 1$ and for every $k \geqslant 1$

$$\frac{C_n}{C_{n+1}} = \frac{1}{4} + \frac{3}{8n-4}, \tag{9}$$

$$\frac{C_n}{C_{n+k}} > \frac{1}{4^k}, \tag{10}$$

$$\lim_{n\to\infty} \frac{C_n}{C_{n+k}} = \frac{1}{4^k}. \tag{11}$$

Definition 10. By $F_n^k(p)$ we mean the number of formulas of size n having p premises, i.e. formulas which are of the form: $\tau = \tau_1 \to (\ldots \to (\tau_p \to \alpha))$, where α is a propositional variable.

Lemma 11. $F_n^k = \sum_{p=1}^{n-1} F_n^k(p)$.

Proof. Since numbers $F_n^k(p)$ are the cardinalities of disjoint sets of formulas for different p's and since there are no formulas of size n having more then $n-1$ premises, for $n \geqslant 2$ we have: $F_n^k = F_n^k(1) + \cdots + F_n^k(n-1)$. □

Definition 12. By $C_n(p)$ we mean $F_n^1(p)$.

As in Lemma 9 we have $F_n^k(p) = k^n C_n(p)$.

Lemma 13. *Number $F_n^k(p)$ is given by the following recursion on p*:

$$F_n^k(0) = \begin{cases} k & \text{if } n = 1, \\ 0 & \text{if } n \neq 1, \end{cases} \tag{12}$$

$$F_n^k(1) = \begin{cases} 0 & \text{if } n = 0, \\ kF_{n-1}^k & \text{if } n > 0, \end{cases} \tag{13}$$

$$F_n^k(p) = \sum_{i=1}^{n-p} F_i^k F_{n-i}^k(p-1). \tag{14}$$

Proof. The formula for $F_n^k(0)$ is obvious. Except for $n = 0$ the number $F_n^k(1) = kF_{n-1}^k$, since $F_n^k(1)$ is the number of formulas of the form $\tau \to \alpha$. There are F_{n-1}^k formulas τ and k propositional variables α. For $p > 1$ consider formula

$$\tau = \tau_1 \to (\underbrace{\tau_2 \to (\ldots(\tau_p \to \alpha)\ldots)}_{\mu}),$$

where τ_1 is of size i. The number of possible formulas of τ is the number of formulas of τ_1 (i.e. F_i^k) and μ (i.e. $F_{n-i}(p-1)$), summed over all possible divisions at position i. The summation stops at $i = n-p$, since beginning with $i = n-p+1$ the terms become zero. □

We are going to isolate the class of simple tautologies which are an important and large fragment of the set of tautologies. As we will see afterwards the class of simple tautologies is so big as to be a good approximation of the whole set of tautologies. Therefore investigations about behavior of the whole set can be nicely approximated by this fragment.

Definition 14. A *simple tautology* is a formula $\tau \in \mathcal{F}_k$ of the form $\tau = \tau_1, \ldots, \tau_p \to \alpha$ such that there is at least one component τ_i identical to α.

Evidently, a simple tautology is a tautology. Let G_n^k be the number of simple tautologies of size n built with k propositional variables and $G_n^k(p)$ be the number of simple tautologies of size n built with k variables with p premises. Our goal is to find how big asymptotically is the fragment of simple tautologies within the set of all formulas.

Lemma 15. *The number G_n^k of simple tautologies is given the recursion*

$$G_1^k = 0, \tag{15}$$

$$G_2^k = k, \tag{16}$$

$$G_n^k = F_{n-1}^k - G_{n-1}^k + \sum_{i=2}^{n-1} F_{n-i}^k G_i^k. \tag{17}$$

Proof. For base cases when $n = 1$ and 2 the proof is trivial. The recursive case is based on two observations: First, $\tau_1 \to \tau_2$ is simple if τ_2 is simple. So for every formula τ_1 of size $n - i$ and every simple tautology τ_2 of size i we have one simple tautology $\tau_1 \to \tau_2$ of size n. The sum starts from $i = 2$ because there are no simple tautologies of size 1. This part is responsible for the component $\sum_{i=2}^{n-1} F^k_{n-i} G^k_i$. The only other simple tautologies are those for which τ_1 is a propositional variable identical to the propositional variable the formula τ_2 points to. Therefore for every formula τ_2 of size $n - 1$ which is not a simple tautology (there are exactly $F^k_{n-1} - G^k_{n-1}$ such formulas) we have exactly one simple tautology $\alpha \to \tau_2$ where α is a propositional variable the formula τ_2 proves. Notice that if τ_2 is already a simple tautology this case is covered by the previous component. □

Lemma 16. *The number $G^k_n(p)$ of simple tautologies with p premises is given by the following recursion on p*:

$$G^k_n(0) = \begin{cases} k & \text{if } n = 1, \\ 0 & \text{if } n \neq 1, \end{cases} \tag{18}$$

$$G^k_n(p+1) = \begin{cases} 0 & \text{if } n \leqslant p, \\ F^k_{n-1}(p) - G^k_{n-1}(p) + \sum_{i=2}^{n-1} F^k_{n-i} G^k_i(p) & \text{if } n > p. \end{cases} \tag{19}$$

Proof. The similar argument as for Lemma 15. Proof must be accompanied with counting the number of premises of the considered simple tautology. □

3. Generating functions

In this paper we investigate the proportion between the number of formulas of the size n that are tautologies against the number of all formulas of size n for propositional formulas of the language $\mathcal{F}_k$. Our interest lies in finding limit of that fraction when $n \to \infty$. For this purpose combinatorics has developed an extremely powerful tool, in the form of generating series and generating functions. A nice exposition of the method can be found in Wilf [7], Comtet [1] as well as in Flajolet, Sedgewick.[3] As the reader may now expect, while working with propositional logic we will be often concerned with complex analysis, analytic functions and their singularities.

Let $A = (A_0, A_1, A_2, \ldots)$ be a sequence of real numbers. The *ordinary generating series* for A is the formal power series $\sum_{n=0}^{\infty} A_n z^n$. And, of course, formal power series are in one-to-one correspondence to sequences. However, considering z as a complex variable, this series, as known from the theory of analytic functions, converges uniformly to a function $f_A(z)$ in some open disc $\{z \in \mathcal{C} : |z| < R\}$ of maximal diameter, and $R \geqslant 0$ is called its radius of convergence. So with the sequence A we can associate a complex function $f_A(z)$, called the *ordinary generating function* for A, defined in a neighborhood of 0. This correspondence is one-to-one again (unless $R = 0$), since, as it is well known from the theory of analytic functions, the expansion of a complex function $f(z)$, analytic in a neighborhood of z_0, into a power series $\sum_{n=0}^{\infty} A_n (z - z_0)^n$ is unique.

Definition 17. Let F be a series in powers of z. Then by the symbol $[z^n]\{F\}$ we will mean the coefficient of z^n in the exponential series expansion of F.

Many questions concerning the asymptotic behavior of A can be efficiently resolved by analyzing the behavior of f_A at the complex circle $|z| = R$. This is the approach we take to determine the asymptotic fraction of tautologies and many other classes of formulas among all formulas of a given size.

Definition 18. The generalized Newton symbol $\binom{a}{n}$ for complex number a stands for $a(a-1)\ldots(a-(n-1))/n!$.

The key tool for finding asymptotics will be the following result, due to Szegö [6, Theorem 8.4], see as well Wilf [7, Theorem 5.3.2, p. 181]. Function v in the assumption of Szegö lemma is the one from which we want to extract coefficients of expansion. Remember that $\xi(q)$ defined in formula (22) is the bound of summation in (21).

[3] P. Flajolet, R. Sedgewick, Analytic combinatorics, symbolic combinatorics, unpublished, see web page http://algo.inria.fr/flajolet/Publications/books.html

Theorem 19 (*Szegö lemma*). *Let $v(z)$ be analytic in $|z| < 1$ with a finite number of singularities $e^{i\varphi^{(k)}}$, $k = 1, \ldots, s$, at the circle $|z| = 1$. Suppose that in the neighborhood of each $e^{i\varphi^{(k)}}$, $v(z)$ has the expansion of the form*

$$v(z) = \sum_{p \geqslant 0} v_p^{(k)} (1 - z e^{-i\varphi^{(k)}})^{a^{(k)} + pb^{(k)}}, \tag{20}$$

where $a^{(k)} \in \mathcal{C}$ and $b^{(k)} > 0$ is real, and the branch chosen above for the expansion equals $v(0)$ for $z = 0$. Then

$$[z^n]\{v(z)\} = \sum_{k=1}^{s} \sum_{p=0}^{\xi(q)} v_p^{(k)} \binom{a^{(k)} + pb^{(k)}}{n} (-e^{i\varphi^{(k)}})^n + O(n^{-q}), \tag{21}$$

with

$$\xi(q) = \max_{k=1\ldots s} \lceil (1/b^{(k)})(q - \Re(a^{(k)}) - 1) \rceil. \tag{22}$$

In all our applications we will indeed have only one singulary. The following is much simpler version of the Szegö Lemma derived from Theorem 19. In this special case we assume to have only one singularity located at $z = 1$. Therefore $s = 1$, $\varphi^{(1)} = 0$. Additionally we assume $b^{(1)} = 1/2$, $a^{(1)} = 0$. Also we will be satisfied with error bound $O(n^{-2})$. So $q = 2$. It follows that our $\xi(q) = 2$. Consult also the paper of Moczurad et al. [4, Corollary 2.4, p. 578]. Under all those assumptions Szegö lemma reduces to the following:

Corollary 20 (*simplified Szegö Lemma*). *Let $v(z)$ be analytic in $|z| < 1$ with $z = 1$ the only singularity at the circle $|z| = 1$. If $v(z)$ in the vicinity of $z = 1$ has the expansion of the form*

$$v(z) = \sum_{p \geqslant 0} v_p (1 - z)^{p/2}, \tag{23}$$

where $p > 0$, and the branch chosen above for the expansion equals $v(0)$ for $z = 0$, then

$$[z^n]\{v(z)\} = \left(v_0 \binom{0}{n} + v_1 \binom{1/2}{n} + v_2 \binom{1}{n} \right) (-1)^n + O(n^{-2}). \tag{24}$$

Moreover, remember that $\binom{0}{n} = \binom{1}{n} = 0$ for $n > 1$ then it reduces even more to:

$$[z^n]\{v(z)\} = v_1 \binom{1/2}{n} (-1)^n + O(n^{-2}). \tag{25}$$

Consult also the simplified Szegö Lemma in Zaionc, [8] and in Kostrzycka and Zaionc [3]. For technical reasons we will need to know the rate of grow of the function $\binom{1/2}{n} (-1)^n$ which appears in formula (25).

Lemma 21. *For $n \in \mathbb{N}$ we have $\binom{1/2}{n} (-1)^{n+1} = O(n^{-3/2})$.*

Proof. It can be obtained from (8) by the Stirling approximation formula (see Robbins [5] for details).

In this part of the section we are going to present the method of finding asymptotic densities for the classes of formulas for which the generating functions are already calculated. The main tool used for this purpose is theorem based on simplified Szegö lemma. The following lemma is a main tool for finding limits of the fraction a_n/b_n, when generating functions for sequences a_n and b_n satisfies conditions of simplified Szegö Lemma 20. □

Theorem 22. *Suppose two functions $v(z)$ and $w(z)$ satisfies assumptions of simplified Szegö theorem* (*Corollary* 20) *i.e. both v and w are analytic in $|z| < 1$ with $z = 1$ being the only singularity at the circle $|z| = 1$. Both $v(z)$ and $w(z)$ in the vicinity of $z = 1$ have expansions of the form*

$$v(z) = \sum_{p \geqslant 0} v_p (1 - z)^{p/2}, \tag{26}$$

$$w(z) = \sum_{p \geqslant 0} w_p (1-z)^{p/2}, \tag{27}$$

then the limit of $[z^n]\{v(z)\}/[z^n]\{w(z)\}$ *exists and is given by formula*:

$$\lim_{n\to\infty} \frac{[z^n]\{v(z)\}}{[z^n]\{w(z)\}} = \frac{v_1}{w_1}. \tag{28}$$

Proof. Applying the main formula (25) from simplified version of Szegö Lemma in Corollary 20 and equation from Lemma 21 we get

$$\begin{aligned}\lim_{n\to\infty} \frac{[z^n]\{v(z)\}}{[z^n]\{w(z)\}} &= \lim_{n\to\infty} \frac{v_1 \binom{1/2}{n} (-1)^n + \mathrm{O}(n^{-2})}{w_1 \binom{1/2}{n} (-1)^n + \mathrm{O}(n^{-2})} \\ &= \lim_{n\to\infty} \frac{-v_1 \mathrm{O}(n^{-3/2}) + \mathrm{O}(n^{-2})}{-w_1 \mathrm{O}(n^{-3/2}) + \mathrm{O}(n^{-2})} = \frac{v_1}{w_1}. \qquad \square\end{aligned}$$

From the previous theorem we can derive the technical lemma which will be very useful for finding limits of the proportion between two sequences of known generating functions.

Lemma 23. *Suppose two functions* $v(z)$ *and* $w(z)$ *satisfies assumptions of simplified Szegö theorem* (*Corollary* 20) *i.e. both* v *and* w *are analytic in* $|z| < 1$ *with* $z = 1$ *being the only singularity at the circle* $|z| = 1$. *Both* $v(z)$ *and* $w(z)$ *in the vicinity of* $z = 1$ *have expansions of the form*

$$v(z) = \sum_{p \geqslant 0} v_p (1-z)^{p/2}, \tag{29}$$

$$w(z) = \sum_{p \geqslant 0} w_p (1-z)^{p/2}. \tag{30}$$

Suppose we have functions $\widetilde{v}$ *and* $\widetilde{w}$ *satisfying* $\widetilde{v}(\sqrt{1-z}) = v(z)$ *and* $\widetilde{w}(\sqrt{1-z}) = w(z)$ *then the limit of* $[z^n]\{v(z)\}/[z^n]\{w(z)\}$ *exists and is given by formula*:

$$\lim_{n\to\infty} \frac{[z^n]\{v(z)\}}{[z^n]\{w(z)\}} = \frac{(\widetilde{v})'(0)}{(\widetilde{w})'(0)}. \tag{31}$$

Proof. Simple consequence of Corollary 20. New functions $\widetilde{v}$ and $\widetilde{w}$ have expansions

$$\widetilde{v}(z) = \sum_{p \geqslant 0} v_p z^p, \tag{32}$$

$$\widetilde{w}(z) = \sum_{p \geqslant 0} w_p z^p. \tag{33}$$

Therefore $v_1 = (\widetilde{v})'(0)$ and $w_1 = (\widetilde{w})'(0)$. By Theorem 22 the result presented in formula 31 is obvious. $\square$

4. Calculating generating functions

We start with calculating generating functions for all recursively defined sequences from Section 2.

Lemma 24. *The generating function* f_F *for the numbers* F_n^k *is*

$$f_F(z) = \frac{1}{2} - \frac{1}{2}\sqrt{1-4kz}. \tag{34}$$

Proof. Obvious. Can be found in paper of Moczurad et al. [4, p. 588] or see the whole exposition in Wilf [7]. As a special case of (34) when $k = 1$ we have generating function f_C for numbers C_n given by $f_C(z) = 1/2 - (\sqrt{1-4z})/2$. □

Lemma 25. *For fixed $p \geqslant 0$ the generating functions $f_{C(p)}$ and $f_{F(p)}$, respectively, for $C_n(p)$ and $F_n^k(p)$ are following*:

$$f_{C(p)}(z) = z \cdot (f_C(z))^p = z \cdot \left(\frac{1-\sqrt{1-4z}}{2}\right)^p, \tag{35}$$

$$f_{F(p)}(z) = k \cdot z \cdot (f_F(z))^p = k \cdot z \cdot \left(\frac{1-\sqrt{1-4kz}}{2}\right)^p. \tag{36}$$

Proof. Let $f_{C(p)}(z)$ be a generating function for $C_n(p)$. Lemma 13 gives $C_n(p) = \sum_{i=1}^{n-p} C_i C_{n-i}(p-1)$ which becomes after a closer examination, the equality $f_{C(p-1)}(z) \cdot f_C(z) = f_{C(p)}(z)$. Since $C_n(1) = C_{n-1}$ we get $f_{C(1)}(z) = z(f_C(z))$ and consequently $f_{C(p)}(z) = z(f_C(z))^p$. Thanks to equation $F_n^k(p) = k^n C_n(p)$ we get $f_{F(p)}(z) = f_{C(p)}(kz)$ which ends the proof of equality (36). Notice that formulas (35) and (36) are also correct for $p = 0$ □

Lemma 26. *The generating function f_G for numbers G_n^k is*

$$f_G(z) = \frac{z f_F(z)}{1 - f_F(z) + z} = \frac{(1+z)(1-\sqrt{1-4\,k\,z}) - 2kz}{2(1+k+z)}. \tag{37}$$

Proof. The recurrence given by Eq. (17)

$$G_n^k = F_{n-1}^k - G_{n-1}^k + \sum_{i=2}^{n-1} F_{n-i}^k G_i^k$$

from Lemma 15 becomes $f_G = f_G \cdot f_F + z \cdot f_F - z \cdot f_G$. Solving it for f_G gives Eq. (37). □

Lemma 27. *For fixed p the generating function $f_{G(p)}$ for $G_n^k(p)$ can be defined by the following recursion on p*:

$$f_{G(0)}(z) = 0,$$
$$f_{G(p+1)}(z) = f_F(z) \cdot f_{G(p)}(z) + kz^2 (f_F(z))^p - z f_{G(p)}(z). \tag{38}$$

Proof. Formula for $f_{G(p+1)}$ is a simple encoding of the recurrence (19). Multiplication $f_F(z) \cdot f_{G(p)}(z)$ is responsible for the fragment $\sum_{i=2}^{n-1} F_{n-i}^k G_i^k(p)$. According to formula (36) (see Lemma 25) for functions $F_n^k(p)$ we have that $kz(f_F(z))^p$ stands for $F_n^k(p)$. Since the number in recurrence depends on $n-1$ not on n it have to be additionally multiply by z. The last fragment $z f_{G(p)}(z)$ is responsible for the recursion $G_{n-1}^k(p)$ in (19). □

As the reader may now expect we are going to prove that for every p function, $f_{G(p)}(z)$ is of the form $C(z) f_F(z) + D(z)$ for certain polynomials $C(z)$ and $D(z)$ of variable z. As we will see also for every p function $(f_F(z))^p$ must have a form of $A(z) f_F(z) + B(z)$ for certain polynomials $A(z)$ and $B(z)$ of variable z.

Definition 28. Let us define four sequences of polynomials by the following mutual recursion:

$$A_0(z) = 0, \quad B_0(z) = 1, \quad C_0(z) = 0, \quad D_0(z) = 0, \tag{39}$$
$$A_{p+1}(z) = A_p(z) + B_p(z), \tag{40}$$
$$B_{p+1}(z) = -kz A_p(z), \tag{41}$$
$$C_{p+1}(z) = C_p(z) + D_p(z) + kz^2 A_p(z) - z C_p(z), \tag{42}$$
$$D_{p+1}(z) = kz^2 B_p(z) - z D_p(z) - kz C_p(z). \tag{43}$$

Theorem 29. *For every $p \geqslant 0$ the following hold*:

$$(f_F(z))^p = A_p(z) f_F(z) + B_p(z), \tag{44}$$

$$f_{G(p)}(z) = C_p(z) f_F(z) + D_p(z), \tag{45}$$

for polynomials $A_p(z)$, $B_p(z)$, $C_p(z)$ and $D_p(z)$ defined recursively in (28).

Proof. Induction on p. For $p = 0$ it is obvious. Since $(f_F(z))^0 = 1$ and $f_{G(0)}(z) = 0$ polynomials are $A(z) = 0$, $B(z) = 1$, $C(z) = 0$ and $D(z) = 0$. The induction step is based on the formula $(f_F(z))^2 = f_F(z) - k \cdot z$ derived from Lemma 24. Suppose $(f_F(z))^p = A_p(z) f_F(z) + B_p(z)$. We can calculate the shape of polynomials $A_{p+1}(z)$ and $B_{p+1}(z)$ in the following way:

$$\begin{aligned}(f_F(z))^{p+1} &= (f_F(z))^p f_F(z)\\ &= (A_p(z) f_F(z) + B_p(z)) f_F(z)\\ &= A_p(z)(f_F(z))^2 + B_p(z) f_F(z)\\ &= A_p(z)(f_F(z) - kz) + B_p(z) f_F(z)\\ &= (A_p(z) + B_p(z)) f_F(z) - kz A_p(z).\end{aligned}$$

Therefore $A_{p+1}(z) = A_p(z) + B_p(z)$ and $B_{p+1}(z) = -kz A_p(z)$. Similarly we calculate $f_{G(p+1)}(z)$ using formula (38) from Lemma 27. Suppose $f_{G(p)}(z) = C_p(z) f_F(z) + D_p(z)$. Calculation is based again on the formula $(f_F(z))^2 = f_F(z) - k \cdot z$ derived from Lemma 24 and on previous formula (44) for $(f_F(z))^p$.

$$\begin{aligned}f_{G(p+1)}(z) &= [f_F(z) - z] \cdot f_{G(p)}(z) + kz^2 (f_F(z))^p\\ &= [f_F(z) - z] \cdot [C_p(z) f_F(z) + D_p(z)] + kz^2 [A_p(z) f_F(z) + B_p(z)]\\ &= C_p(z)[f_F(z)]^2 + [D_p(z) + kz^2 A_p(z) - zC_p(z)] f_F(z) + kz^2 B_p(z) - zD_p(z)\\ &= C_p(z)[f_F(z) - kz] + [D_p(z) + kz^2 A_p(z) - zC_p(z)] f_F(z) + kz^2 B_p(z) - zD_p(z)\\ &= [C_p(z) + D_p(z) + kz^2 A_p(z) - zC_p(z)] f_F(z) kz^2 B_p(z) - zD_p(z) - kzC_p(z).\end{aligned}$$

Therefore new polynomial coefficients for $C_{p+1}(z)$ and $D_{p+1}(z)$ are expressible by the old ones in the way described above. □

The first few generation functions $f_{G(p)}(z)$ are the following:

$$\begin{aligned}f_{G(1)}(z) &= kz^2,\\ f_{G(2)}(z) &= 2kz^2 f_F(z) - kz^3,\\ f_{G(3)}(z) &= (3kz^2 - 3kz^3) f_F(z) - 3k^2 z^3 + kz^4,\\ f_{G(4)}(z) &= (4kz^2 - 6kz^3 - 4k^2 z^3 + 4kz^4) f_F(z) - 4k^2 z^3 + 6k^2 z^4 - kz^5.\end{aligned}$$

Expanding $f_F(z)$ in above formulas gives the combinatorial insight of the number of simple tautologies with the certain number of premises. For example in the expansion of $f_{G(2)}(z)$ we can see the pattern e.g.

$$f_{G(2)}(z) = (2k^2 - k)z^3 + 2k^3 z^4 + 4k^4 z^5 + 10k^5 z^6 + 28k^6 z^7 + 84k^7 z^8 + \cdots$$

The natural combinatorial binary tree interpretation is that the number of simple tautologies for $n \geqslant 4$ with two premises is twice as many as the total number of formulas shorter by 2 multiplied by the number k of labels.

5. Calculation of limits

In this section we are going to find asymptotic densities for the classes of formulas for which the generating functions are already calculated. The main tool used for this purpose is Lemma 23. The main goal of this section is to find the formula for the asymptotic density of the classes of simple tautologies with p premises which later on allows us to speak about distribution of probabilities.

First we recall two results from Moczurad et al. [4]. In the first one we consider the probability that the given formula is simple tautology. The meaning of this theorem is that the limit of the fraction G_n^k / F_n^k while n tends to infinity exists

and the size of true formulas is at least as big as $O(1/k)$. In fact in paper [4] we proved that the size of true formulas is also at most as big as $O(1/k)$ (see [4, Corollary 6.10, p. 587]). The second theorem finds the probability that the given formula has p premises. Both are good examples of the usefulness of Theorem 22 and Lemma 23.

Theorem 30.

$$\lim_{n\to\infty} \frac{G_n^k}{F_n^k} = \frac{4k+1}{(2k+1)^2}. \tag{46}$$

Proof. We show now much more efficient proof based on Theorem 22 and Lemma 23 compared with those from paper [4]. Indeed, first we recall Eq. (37) from Lemma 26 for f_G and formula for f_F from Lemma 24.

$$f_G(z) = \frac{(1+z)(1-\sqrt{1-4k\,z}) - 2kz}{2(1+k+z)},$$
$$f_F(z) = \frac{1}{2} - \frac{1}{2}\sqrt{1-4kz}.$$

In order to satisfy assumptions of Theorem 22 we normalize functions in such a way to have the only singularity located in $|z| \leqslant 1$ at the position in $z = 1$. So, we define functions $\overline{f_G}(z) = f_G(z/(4k))$ and $\overline{f_F}(z) = f_F(z/(4k))$. Therefore we have

$$\overline{f_G}(z) = -\frac{1}{2}\frac{-z-4k+2kz+(4k+z)\sqrt{1-z}}{4k(1+k)+z},$$
$$\overline{f_F}(z) = \frac{1}{2} - \frac{1}{2}\sqrt{1-z}. \tag{47}$$

This representation reveals that the only singularity of $\overline{f_G}(z)$ and $\overline{f_F}(z)$ located in $|z| \leqslant 1$ is indeed $z = 1$. We have to remember that change of a caliber of the radius of convergence for functions f_G and f_F effects accordingly sequences represented by the new functions. Therefore we have $G_n^k = (4k)^n([z^n]\{\overline{f_G}(z)\})$ and $F_n^k = (4k)^n([z^n]\{\overline{f_F}(z)\})$. Now we are ready to use Lemma 23. Let us define functions $\widetilde{f_F}$ and $\widetilde{f_G}$ so as to satisfy the following equations: $\widetilde{f_F}(\sqrt{1-z}) = \overline{f_F}(z)$ and $\widetilde{f_G}(\sqrt{1-z}) = \overline{f_G}(z)$. Functions $\widetilde{f_F}$ and $\widetilde{f_G}$ are defined in the following way:

$$\widetilde{f_G}(z) = -\frac{1}{2}\frac{(z-1)^2}{z-2k-1}, \tag{48}$$
$$\widetilde{f_F}(z) = \frac{1}{2} - \frac{1}{2}z. \tag{49}$$

The derivatives $(\widetilde{f_F})'$ and $(\widetilde{f_F})'$ are the following:

$$(\widetilde{f_G})'(z) = -\frac{1}{2}\frac{(z-4k-1)(z-1)}{(z-2k-1)^2},$$
$$(\widetilde{f_F})'(z) = -\frac{1}{2}.$$

Finally derivatives $(\widetilde{f_F})'$ and $(\widetilde{f_F})'$ at $z = 0$ are:

$$(\widetilde{f_G})'(0) = -\frac{1}{2}\frac{4k+1}{(2k+1)^2},$$
$$(\widetilde{f_F})'(0) = -\frac{1}{2}.$$

Now applying Lemma 23 we get

$$\lim_{n\to\infty} \frac{G_n^k}{F_n^k} = \lim_{n\to\infty} \frac{(4k)^n\left([z^n]\{\overline{f_G}(z)\}\right)}{(4k)^n\left([z^n]\{\overline{f_F}(z)\}\right)} = \frac{(\widetilde{f_G})'(0)}{(\widetilde{f_F})'(0)} = \frac{4k+1}{(2k+1)^2},$$

which ends the proof. □

The proof of Theorem 30 reveals the "technology" of determining the convergence of fractions in which both numerator and denominator are given recursively and both generating functions are satisfying simplified Szegö Lemma 23. The proof of the next theorem will use exactly the same technique.

Lemma 31. *The asymptotic probability of the fact that a random formula admits exactly p premises is*

$$\lim_{n\to\infty} \frac{F_n^k(p)}{F_n^k} = \frac{p}{2^{p+1}}. \tag{50}$$

Proof. First we recall Eq. (36) from Lemma 25 describing function $f_{F(p)}$. All steps for denominator f_F are already done in previous Theorem 30.

$$f_{F(p)}(z) = kz(f_F(z))^p = kz\left(\frac{1-\sqrt{1-4kz}}{2}\right)^p. \tag{51}$$

Function $\overline{f_{F(p)}}(z) = f_{F(p)}(z/(4k))$ defined to satisfy Theorem 22 is as follows:

$$\overline{f_{F(p)}}(z) = \frac{z}{4}(\overline{f_F}(z))^p = \frac{z}{4}\left(\frac{1-\sqrt{1-z}}{2}\right)^p. \tag{52}$$

It is clear that $\overline{f_{F(p)}}(z)$ admits the only singularity at $z = 1$. As in previous theorem let us define functions $\widetilde{f_{F(p)}}$ as to satisfy the following equations: $\widetilde{f_{F(p)}}(\sqrt{1-z}) = \overline{f_F(p)}(z)$. Therefore

$$\widetilde{f_{F(p)}}(z) = \frac{1-z^2}{4}\left(\frac{1-z}{2}\right)^p. \tag{53}$$

Derivative of the function $\widetilde{f_{F(p)}}(z)$ is following:

$$(\widetilde{f_{F(p)}})'(z) = -\frac{z^2}{2}\left(\frac{1-z}{2}\right)^2 - p\frac{(1-z^2)}{8}\left(\frac{1-z}{2}\right)^{p-1}. \tag{54}$$

Therefore $(\widetilde{f_{F(p)}})'(0) = -\frac{1}{2}p/(2^{p+1})$ which concludes the proof. □

The main goal of this section is to find the formula for the asymptotic density of the classes of simple tautologies with p premises which allows us to speak about distribution of probabilities. This part is based on the Theorem 29 which shows very specific form of each function from two families of $f_{G(p)}(z)$ and $(f_F(z))^p$ for all $p \geqslant 0$. This will be a starting point for the construction of the recursive definition of the limit of each function in terms of the previous limits.

Lemma 32. *Let $h(z) = A(z)f_F(z)$ be a generating function for some sequence H_n where $A(z)$ is some polynomial of variable z. The sequence of fractions H_n/F_n^k admits limit.*

Proof. We can easily observe the existence of limit for function $A(z) = z^s$ for $s \geqslant 0$. Function $z^s f_F(z)$ is a generating function for the sequence with the limit property. It is due to formula (11). We get

$$\lim_{n\to\infty} \frac{H_n}{F_n^k} = \lim_{n\to\infty} \frac{F_{n-s}^k}{F_n^k} = \lim_{n\to\infty} \frac{k^{n-s}C_{n-s}}{k^n C_n} = \lim_{n\to\infty} \frac{1}{k^s}\frac{C_{n-s}}{C_n} = \frac{1}{(4k)^s}.$$

Because of linear property of limits we have got the limit for all functions on the form $h(z) = A(z)f_F(z)$. If a polynomial $A(z) = \sum_{i=0}^k A_i z^i$ then the limit is $\sum_{i=0}^k A_i/(4k)^i$. □

Lemma 33. *Let $g(z) = A(z)f_F(z) + B(z)$ where $A(z)$ and $B(z)$ are some polynomials of variable z. Let $h(z) = A(z)f_F(z)$ be generating functions for some sequences G_n and H_n, respectively. Two sequences of fractions H_n/F_n^k and G_n/F_n^k admit limits and limits are identical.*

Proof. Trivially follows from Lemma 32. A function in the form $g(z) = A(z) f_F(z) + B(z)$ admits the identical limit $\sum_{i=0}^{k} A_i/(4k)^i$ since the coefficients of $B(z)$ disturb only first few finite numbers of coefficients from an expansion of the $A(z) f_F(z)$ but this does not make any effect on the limit. □

Now we are ready to define recursive dependencies between limits of sequences associated with the classes of four different polynomials $A_p(z)$, $B_p(z)$, $C_p(z)$ and $D_p(z)$ defined in Definition 28.

Definition 34. Let A be a polynomial. Let $h(z) = A(z) f_F(z)$ be a generating function for some sequence H_n. According to Lemma 32 the sequence of fractions H_n/F_n^k admits limit. By $\overrightarrow{A}$ we mean the limit of this sequence, so

$$\overrightarrow{A} = \lim_{n\to\infty} \frac{H_n}{F_n^k} = \lim_{n\to\infty} \frac{[z^n]\{A(z) f_F(z)\}}{F_n^k}.$$

Lemma 35. *For any polynomials A and B and a number α the limits have to satisfy the following obvious conditions*:

$$\overrightarrow{A+B} = \overrightarrow{A} + \overrightarrow{B}, \tag{55}$$

$$\overrightarrow{\alpha A} = \alpha \overrightarrow{A}, \tag{56}$$

$$\overrightarrow{zA(z)} = \frac{1}{4k} \overrightarrow{A}. \tag{57}$$

Proof. Eqs. (55) and (56) are obvious. For (57) observe the calculation of limit in Lemma 32. □

Our goal now is to establish recursive dependencies between sequences of limits $\overrightarrow{A_p}$, $\overrightarrow{B_p}$, $\overrightarrow{C_p}$, and $\overrightarrow{D_p}$. Four sequences of polynomials A_p, B_p, C_p and D_p are defined by mutual recursion in Definition 28. The recurrence between limits can be found straightforwardly by encoding the definitions of the polynomials itself. We are specially interested in the family of limits $\overrightarrow{C_p}$, since for the given p it is in fact an asymptotic probability of the class of simple tautologies with p premises. We start with independent solution for limits $\overrightarrow{A_p}$ and $\overrightarrow{B_p}$.

Lemma 36. $\overrightarrow{A_p} = \frac{p}{2^{p-1}}$.

Proof. Sequence of limits $\overrightarrow{A_p}$ satisfies the following recursive definition:

$$\overrightarrow{A_0} = 0, \quad \overrightarrow{A_1} = 1, \tag{58}$$

$$\overrightarrow{A_{p+1}} = \overrightarrow{A_p} - \frac{1}{4}\overrightarrow{A_{p-1}}. \tag{59}$$

since from Lemma 28 we can find easily the independent recurrence on polynomials A_p namely $A_{p+1}(z) = A_p(z) - kzA_{p-1}(z)$. Formula (59) is due to the simple computation on limits using rules (55)–(57). Now it is straightforward to solve the three term "Fibonacci like" recurrence (see for example Wilf [7, p. 8]). Generating function $\mathcal{A}(z)$ for the sequence of limits $\overrightarrow{A_p}$ have to satisfy an equation $(\mathcal{A}(z) - z)/z = \mathcal{A}(z) - \frac{1}{4}z\mathcal{A}(z)$. Solving it gives $\mathcal{A}(z) = 4z/(z-2)^2$. Since the function $z/(az+b)^2$ has the expansion $\sum_{i=0}^{\infty} -i/ab(-a/b)^i z^i$, we get the solution. □

Lemma 37. $\overrightarrow{B_p} = -\frac{p-1}{2^p}$.

Proof. Immediate from $B_{p+1}(z) = -kzA_p(z)$. □

Theorem 38. $\overrightarrow{C_p} = \frac{p}{2^{p+1}} - p\frac{(2k-1)^{p-1}}{4^p k^{p-1}}$.

Proof. Sequences of limits $\overrightarrow{C_p}$ and $\overrightarrow{D_p}$ have to satisfy the following mutual recursive definition:

$$\overrightarrow{C_{p+1}} = \left(1 - \frac{1}{4k}\right)\overrightarrow{C_p} + \overrightarrow{D_p} + \frac{p}{k \cdot 2^{p+3}}, \tag{60}$$

$$\overrightarrow{D_{p+1}} = -\frac{\overrightarrow{C_p}}{4} - \frac{\overrightarrow{D_p}}{4k} - \frac{(p-1)}{k \cdot 2^{p+4}}. \tag{61}$$

This is immediate from (42) and (43) of Definition 28 and previous Lemmas 36 and 37. Now we solve recursive equations (60) and (61) in the standard way by creating appropriate generation functions for two sequences of limits $\overrightarrow{C_p}$ and $\overrightarrow{D_p}$. We start with finding the equations between the generating functions which describe the recurrence (60) and (61). Let $\mathcal{C}$ and $\mathcal{D}$ be generating functions for sequences of limits $\overrightarrow{C_p}$ and $\overrightarrow{D_p}$, respectively. Functions $\mathcal{C}$ and $\mathcal{D}$ satisfy the following equations:

$$\frac{\mathcal{C}(z)}{z} = \left(1 - \frac{1}{4k}\right)\mathcal{C}(z) + \mathcal{D}(z) + \frac{z}{4k(2-z)^2}, \tag{62}$$

$$\frac{\mathcal{D}(z)}{z} = -\frac{\mathcal{C}(z)}{4} - \frac{\mathcal{D}(z)}{4k} - \frac{(z-1)}{4k \cdot (1-z)^2}. \tag{63}$$

To find equation for $\mathcal{C}$ multiply both sides of recurrence relation (60)

$$\overrightarrow{C_{p+1}} = \left(1 - \frac{1}{4k}\right)\overrightarrow{C_p} + \overrightarrow{D_p} + \frac{p}{k \cdot 2^{p+3}} \tag{64}$$

by z^p and sum over the values on p for which the recurrence is valid namely for $p \geqslant 0$. On left side we get $\sum_{p=0}^{\infty} \overrightarrow{C_{p+1}} z^p$. It is the same with $(\sum_{p=0}^{\infty} \overrightarrow{C_p} z^p - \overrightarrow{C_0})/z$ which is $\mathcal{C}(z)/z$ (since $\overrightarrow{C_0} = 0$). On the right side it is immediate $\sum_{p=0}^{\infty}(1 - 1/4k)\overrightarrow{C_p} z^p = (1-1/4k)\mathcal{C}(z)$ and $\sum_{p=0}^{\infty} \overrightarrow{D_p} z^p = \mathcal{D}(z)$. The last fragment $\sum_{p=0}^{\infty} p/(k \cdot 2^{p+3}) z^p = 1/8k \sum_{p=0}^{\infty} p(\frac{1}{2})^p$ which after summation becomes $z/4k(2-z)^2$. The similar summation we do for the recurrence relation.

$$\overrightarrow{D_{p+1}} = -\frac{\overrightarrow{C_p}}{4} - \frac{\overrightarrow{D_p}}{4k} - \frac{(p-1)}{k \cdot 2^{p+4}} \tag{65}$$

to obtain

$$\frac{\mathcal{D}(z)}{z} = -\frac{\mathcal{C}(z)}{4} - \frac{\mathcal{D}(z)}{4k} - \frac{(z-1)}{4k \cdot (1-z)^2}.$$

The only solution for $\mathcal{C}(z)$ presented in the form of partial fractions is as follows:

$$\mathcal{C}(z) = \frac{2}{(z-2)^2} + \frac{1}{(z-2)} - \frac{16k^3}{(2k-1)((2k-1)z-4k)} - \frac{4k^2}{(2k-1)((2k-1)z-4k)^2}. \tag{66}$$

We use the standard expansion formulas separately for every fraction in (66). To conclude the proof we extract pth element of expansion from every formula and we get

$$\overrightarrow{C_p} = \frac{p}{2^{p+1}} - p\frac{(2k-1)^{p-1}}{4^p k^{p-1}} \qquad \square$$

We have found the simple formula

$$\frac{p}{2^{p+1}} - p\frac{(2k-1)^{p-1}}{4^p k^{p-1}}$$

for $\overrightarrow{C_p}$. The natural and intended interpretation of $\overrightarrow{C_p}$ is the probability that the random implicational formula is a simple tautology with p premises.

Theorem 39. *The asymptotic probability of the fact that a random formula is a simple tautology with exactly p premises is*

$$\lim_{n\to\infty} \frac{G_n^k(p)}{F_n^k} = \frac{p}{2^{p+1}} - p\frac{(2k-1)^{p-1}}{4^p k^{p-1}}. \tag{67}$$

Proof. Base consequently on Theorem 29, Lemma 33 and finally Theorem 38,

$$\begin{aligned}\lim_{n\to\infty}\frac{G_n^k(p)}{F_n^k} &= \lim_{n\to\infty}\frac{[z^n](f_{G(p)}(z))}{[z^n](f_F(z))}\\ &= \lim_{n\to\infty}\frac{[z^n](C_p(z)f_F(z)+D_p(z))}{[z^n](f_F(z))}\\ &= \lim_{n\to\infty}\frac{[z^n](C_p(z)f_F(z))}{[z^n](f_F(z))} = \overrightarrow{C_p}\\ &= \frac{p}{2^{p+1}} - p\frac{(2k-1)^{p-1}}{4^p k^{p-1}}. \qquad \square\end{aligned}$$

Theorem 40. *The probability that simple tautology has exactly p premises is described by*

$$\lim_{n\to\infty}\frac{G_n^k(p)}{G_n^k} = \left(\frac{(2k+1)^2}{4k+1}\right)\left(\frac{p}{2^{p+1}} - p\frac{(2k-1)^{p-1}}{4^p k^{p-1}}\right). \tag{68}$$

Proof. Combine two limit equations from Theorems 30 and 39. $\square$

6. Distribution of probabilities

In this section we will discuss and compare the distribution of probabilities proved in previous sections. There are two main questions we wish to discuss:

What is the probability that a randomly chosen implicational formula admits p premises?

What is the probability that a randomly chosen implicational simple tautology admits p premises?

To answer the first question we group together all formulas with p premises and according to Definition 1 we try to find the asymptotic probability of this class. But this is exactly what we have found in Theorem 31. So let us start with analyzing the first distribution:

Definition 41. Let us define the random variable X which assigns to a implicational formula the number of its premises.

Lemma 42. *Random variable X has the distribution* $\overline{\overline{X}}(p) = \lim_{n\to\infty}(F_n^k(p)/F_n^k) = p/2^{p+1}$, *expected value* $E(\overline{X}) = 3$, *variance* $Var(\overline{X}) = 4$. *The standard deviation of X is* 2.

Proof. Technical observation. As we know the number of formulas of size n with the p premises is $F_n^k(p)$. Therefore according to Lemma 31 the asymptotic probability is $p/2^{p+1}$. This forms a distribution since $\sum_{p=0}^{\infty} p/2^{p+1} = 1$. Expected value $E(\overline{X}) = \sum_{p=1}^{\infty} p\ \overline{X}(p) = \sum_{p=1}^{\infty} p(p/2^{p+1}) = 3$, and variance $Var(\overline{X}) = E(\overline{(X - E(X))^2}) = E(\overline{X^2}) - (E(\overline{X}))^2 = \sum_{p=1}^{\infty} p^2(p/2^{p+1}) - 9 = 4$, so the standard deviation of X is $\sqrt{Var(\overline{X})} = 2$. $\square$

As the trivial consequences of the lemma above we can see that surprisingly typical implicational formula have exactly 3 premises. Consider the set of formulas

$$\left\{\phi : |X(\phi) - E(\overline{X})| \leqslant \sqrt{Var(\overline{X})}\right\}.$$

The asymptotic density of this set of formulas with premises laying between 1 and 5 is asymptotically fairly big and amounts to $\sum_{p=1}^{5} p^2/2^{p+1} = 57/64$ which is about 89%.

Now we will start to answer the second question. First we have to isolate the class of all simple tautologies with p premises and compare it against the class of all simple tautologies. But this is exactly what we have found in Theorem 40. We will see now the difference between distribution of the number of premises for all formulas contrasted with the same distribution for simple tautologies only.

Definition 43. For every $k \geqslant 1$ separately let us define the random variable Y_k which assigns to a implicational simple tautology in the language $\mathcal{F}_k$ the number of its premises.

Theorem 44. *Random variable Y_k has the following distribution, expected value and variance*:

$$\overline{Y_k}(p) = \lim_{n\to\infty} \frac{G_n^k(p)}{G_n^k} = \left(\frac{(2k+1)^2}{4k+1}\right)\left(\frac{p}{2^{p+1}} - p\frac{(2k-1)^{p-1}}{4^p k^{p-1}}\right),$$

$$E(\overline{Y_k}) = \frac{40k^2+18k+3}{(2k+1)(4k+1)},$$

$$Var(\overline{Y_k}) = \frac{384k^4+288k^3+160k^2+48k+4}{(2k+1)^2(4k+1)^2}.$$

Proof. As we know, the number of simple tautologies with p premises is $G_n^k(p)$. The asymptotic probability $\lim_{n\to\infty} G_n^k(p)/G_n^k$ is computed in Theorem 40. This constitutes a distribution since $\sum_{p=0}^{\infty} \overline{Y_k}(p) = 1$ (for summation use formula $\sum_{i=0}^{\infty} i z^i = z/(1-z)^2$ twice). Expected value of Y_k (for summation use formula $\sum_{i=0}^{\infty} i^2 z^i = z(1+z)/(1-z)^3$ twice) is $E(\overline{Y_k}) = \sum_{p=0}^{\infty} p\overline{Y_k}(p) = (40k^2+18k+3)/(2k+1)(4k+1)$. Comparing this with the distribution $\overline{X}(p)$ reader can easily check that starting with $k = 1$ the expected value of the number of premises for simple tautologies is substantially greater then 3 and is growing asymptotically to 5 since $\lim_{k\to\infty} E(\overline{Y_k}) = 5$. Variance (use formula $\sum_{i=0}^{\infty} i^3 z^i = z(1+4z+z^2)/(1-z)^4$) is $Var(\overline{Y_k}) = E(\overline{(Y_k - E(Y_k))^2}) = E(\overline{(Y_k)^2}) - (E(\overline{Y_k}))^2$ $= (384k^4+288k^3+160k^2+48k+4)/(2k+1)^2(4k+1)^2$. Asymptotic behavior of $Var(\overline{Y_k})$ can be easily found as $\lim_{k\to\infty} Var(\overline{Y_k}) = 6$. □

7. Limit distribution

The natural question is how the distribution of true sentences look like for very large numbers k and does there exist an uniform asymptotic distribution when k, the number of propositional variables in the logic, tends to infinity. The answers are following:

Lemma 45. *For fixed $p \geqslant 0$*

$$\lim_{k\to\infty} \overline{Y_k}(p) = \frac{p(p-1)}{2^{p+2}}. \tag{69}$$

Proof. See the formula for $\overline{Y_k}(p)$ from Theorem 44. For $p = 0$ and 1 it is obvious. For $p \geqslant 2$ it is simple limit exercise:

$$\begin{aligned}
&\lim_{k\to\infty}\left(\frac{(2k+1)^2}{4k+1}\right)\left(\frac{p}{2^{p+1}} - p\frac{(2k-1)^{p-1}}{4^p k^{p-1}}\right)\\
&\quad= \lim_{k\to\infty}\left(\frac{(2k+1)^2}{k(4k+1)}\frac{p}{2^{p+1}}\right)\left(k - \frac{2(2k-1)^{p-1}}{2^p k^{p-2}}\right)\\
&\quad= \lim_{k\to\infty}\left(\frac{(2k+1)^2}{k(4k+1)}\frac{p}{2^{p+1}}\right)\left(\frac{2\binom{p-1}{1}(2k)^{p-2} + 2\binom{p-1}{2}(2k)^{p-3} - \cdots}{4(2k)^{p-2}}\right)\\
&\quad= \lim_{k\to\infty}\left(\frac{(2k+1)^2}{k(4k+1)}\frac{p}{2^{p+1}}\right)\left(\frac{p-1}{2} + \frac{(p-1)(p-2)}{4(2k)} - \cdots\right)\\
&\quad= \frac{p(p-1)}{2^{p+2}}. \qquad \square
\end{aligned}$$

Definition 46. Let us define the limit distribution $\overline{Y_\infty}$ by $\overline{Y_\infty}(p) = p(p-1)/2^{p+2}$.

This is in fact distribution since $\sum_{p=0}^{\infty} \overline{Y_\infty}(p) = \sum_{p=0}^{\infty} p(p-1)/2^{p+2} = 1$. Expected value of Y_∞ is $E(\overline{Y_\infty}) = \sum_{p=0}^{\infty} p\overline{Y_\infty}(p) = \sum_{p=0}^{\infty} p^2(p-1)/2^{p+2} = 5$. The variance of $\overline{Y_\infty}$ is $Var(\overline{Y_\infty}) = E(\overline{(Y_\infty - E(\overline{Y_\infty}))^2}) = E((\overline{Y_\infty})^2) - (E(\overline{Y_\infty}))^2 = \sum_{p=0}^{\infty} p^2 p(p-1)/2^{p+2} - 25 = 31 - 25 = 6$. So it is clear now that

$$\forall\, p \geqslant 0 \quad \lim_{k\to\infty} \overline{Y_k}(p) = \overline{Y_\infty}(p), \tag{70}$$

$$\lim_{k\to\infty} E(\overline{Y_k}) = E(\overline{Y_\infty}), \tag{71}$$

$$\lim_{k\to\infty} Var(\overline{Y_k}) = Var(\overline{Y_\infty}). \tag{72}$$

The componentwise convergence presented in Lemma 45 and summarized by formula (70) can be extended to much stronger uniform convergence. Below we show the uniformity of convergence of the sequence of distributions $\overline{Y_k}$ to $\overline{Y_\infty}$ when k tends to infinity. Therefore in fact the distribution $\overline{Y_\infty}$ can be treated as a good model of distribution for simple tautologies for the language $\mathcal{F}_k$ when the number k of atomic propositional variables is large.

Theorem 47. *The sequence of distributions $\overline{Y_k}$ uniformly converges to the distribution $\overline{Y_\infty}$.*

Proof. It can be shown by very laborious but simple calculations of the Cartesian distance between distributions $\overline{Y_k}$ and $\overline{Y_\infty}$. The distance between functions $f : \mathbb{N} \to \mathbb{R}$ and $g : \mathbb{N} \to \mathbb{R}$ is defined by

$$dis(f, g) = \sum_{p=0}^{\infty} (f(p) - g(p))^2 .$$

Since we have explicit formulas expressing $\overline{Y_k}$ and $\overline{Y_\infty}$ we are able to find an expression for the distance written only in terms of k.

$$\begin{aligned} dis(\overline{Y_k}, \overline{Y_\infty}) &= \sum_{p=0}^{\infty} (\overline{Y_k}(p) - \overline{Y_\infty}(p))^2 \\ &= \sum_{p=0}^{\infty} ((\overline{Y_k}(p))^2 - 2\sum_{p=0}^{\infty} \overline{Y_\infty}(p)\overline{Y_k}(p) + \sum_{p=0}^{\infty} (\overline{Y_\infty}(p))^2. \end{aligned}$$

Let us calculate separately each sum. Notice that each one is of the form of some combination of known power series $\sum_{i=0}^{\infty} i^s z^i$ for some s.

$$\sum_{p=0}^{\infty} ((\overline{Y_k}(p))^2 = \frac{(2k+1)^4}{(4k+1)^2}\left(\frac{5}{27} + \frac{16k^4(20k^2-4k+1)}{(12k^2+4k-1)^3} - \frac{8k^2(10k-1)}{(6k+1)^3}\right)$$

$$\sum_{p=0}^{\infty} \overline{Y_\infty}(p)\overline{Y_k}(p) = \frac{(2k+1)^2}{(4k+1)}\left(\frac{1}{9} - \frac{k}{4(2k-1)}\left(\frac{16k(2k-1)^2(18k-1)}{(6k+1)^4}\right)\right).$$

$$\sum_{p=0}^{\infty} (\overline{Y_\infty}(p))^2 = \frac{11}{81}.$$

So altogether,

$$\begin{aligned} dis(\overline{Y_k}, \overline{Y_\infty}) = {} & \frac{-2936}{3375\,(1+4\,k)^2} + \frac{126656}{16875\,(1+4\,k)} + \frac{32}{6075\,(6\,k-1)^3} \\ & + \frac{632}{30375(6k-1)^2} + \frac{5144}{151875(6k-1)} + \frac{128}{81(1+6k)^4} \\ & - \frac{1280}{243(1+6k)^3} + \frac{2056}{243(1+6k)^2} - \frac{2744}{243(1+6k)}. \end{aligned}$$

Now, having exact term for $dis(\overline{Y_k}, \overline{Y_\infty})$ presented in the form of partial fractions it is straightforward that $\lim_{k\to\infty} dis(\overline{Y_k}, \overline{Y_\infty}) = 0$. This method is sufficient to show uniform convergence since $(\overline{Y_k}(p) - \overline{Y_\infty}(p))^2 \leqslant dis(\overline{Y_k}, \overline{Y_\infty})$,

for all p. Therefore,

$$|\overline{Y_k}(p) - \overline{Y_\infty}(p)| \leqslant \sqrt{dis(\overline{Y_k}, \overline{Y_\infty})}. \qquad \square \tag{73}$$

Corollary 48. *For fixed $k > 0$ and $p > 0$*

$$\lim_{n\to\infty} \frac{G_n^k(p)}{F_n^k(p)} = 1 - \left(\frac{2k-1}{2k}\right)^{p-1}. \tag{74}$$

Proof. We are going to combine together formula (50) from Lemma 31 with the main result given in formula (67) from Theorem 39. Simple calculation on limits since for $p > 0$

$$\begin{aligned}\lim_{n\to\infty} \frac{G_n^k(p)}{F_n^k(p)} &= \lim_{n\to\infty} \frac{G_n^k(p)}{F_n^k} \cdot \lim_{n\to\infty} \frac{F_n^k}{F_n^k(p)} \\ &= \left(\frac{p}{2^{p+1}} - p\frac{(2k-1)^{p-1}}{4^p k^{p-1}}\right)\left(\frac{2^{p+1}}{p}\right) \\ &= 1 - \left(\frac{2k-1}{2k}\right)^{p-1}. \qquad \square\end{aligned}$$

The result shows how big asymptotically the size of the fraction of simple tautologies with p premises among all formulas of p premises is. We can see that with p growing this fraction becomes closer and closer to 1. Of course, the fraction of all, not only simple, tautologies with p premises is even larger. So the "*density of truth*'' within the classes of formulas of p premises can be as big as we wish. For every $\varepsilon > 0$ we can effectively find p such that among formulas with p premises almost all formulas (except for a tiny fraction of the size ε) asymptotically are tautologies. This should be contrasted with the results proved in Theorem 30. It shows that *density of truth* for all $p's$ together is always of the size $O(1/k)$. The result for every p treated separately is very different. Based on Corollary 48 we may try to estimate the probability for a random long implicational formula to be a tautology by the "*probabilistic algorithm*" algorithm below.

Given: Implicational formula ϕ from $\mathcal{F}_k$.
Problem: Estimate the chances for ϕ to be a tautology.
Solution: Count the number of premises p in ϕ. Then the chances for ϕ to be a tautology are $1 - ((2k-1)/2k)^{p-1}$.

The algorithm above can be performed quickly in terms of the length of formula ϕ. In the worst case we need a linear time to compute the number of premises of ϕ. However the average case time for the algorithm is $O(\log n)$. The accuracy of the algorithm can be estimated using Eq. (73).

Acknowledgements

The part of symbolic computations especially solution of system of Eqs. (62), (63) and the presentation of $\mathcal{C}(z)$ in formula (66) in the form of partial fractions and calculations of the Cartesian distance between distributions presented in the proof of Theorem 47 has been done using *Mathematica*® [4] package. Thanks to the anonymous referee for his valuable comments.

References

[1] L. Comtet, Advanced combinatorics, The Art of Finite and Infinite Expansions, revised and enlarged ed., Reidel, Dordrecht, 1974.
[2] M. Harris, Counting satisfiable k-cnf formulas, Lecture Notes Comput. Sci. 2239 (2001) 765.
[3] Z. Kostrzycka, M. Zaionc, Statistics of intuitionistic versus classical logics, Studia Logica 76 (3) (2004) 307–328.
[4] M. Moczurad, J. Tyszkiewicz, M. Zaionc, Statistical properties of simple types, Math. Struct. Comput. Sci. 10 (2000) 575–594.

[4] *Mathematica* is a registered trademark of Wolfram Research.

[5] H. Robbins, A remark on Stirling's formula, Amer. Math. Monthly 62 (1955) 26–29.
[6] G. Szegö, Orthogonal Polynomials, fourth ed., Vol. 23, AMS, Colloquium Publications, Providence, RI, 1975.
[7] H.S. Wilf, Generatingfunctionology, second ed., Academic Press, Boston, 1994.
[8] M. Zaionc, On the asymptotic density of tautologies in logic of implication and negation, Rep. Math. Logic 39 (2004).